

Modeling High Dimensional Multichannel Brain Signals

Lechuan Hu¹, Norbert Fortin² and Hernando Ombao^{1,3,4}

Abstract

In this paper, our goal is to model functional and effective (directional) connectivity in network of multichannel brain physiological signals (e.g., electroencephalograms, local field potentials). The primary challenges here are twofold: first, there are major statistical and computational difficulties for modeling and analyzing high dimensional multichannel brain signals; second, there is no set of universally-agreed measures for characterizing connectivity.

To model multichannel brain signals, our approach is to fit a vector autoregressive (VAR) model with sufficiently high order so that complex lead-lag temporal dynamics between the channels can be accurately characterized. However, such a model contains a large number of parameters. Thus, we will estimate the high dimensional VAR parameter space by our proposed hybrid LASSLE method (LASSO+LSE) which imposes regularization on the first step (to control for sparsity) and constrained least squares estimation on the second step (to improve bias and mean-squared error of the estimator). Then to characterize connectivity between channels in a brain network, we will use various measures but put an emphasis on partial directed coherence (PDC) in order to capture directional connectivity between channels. PDC is a directed frequency-specific measure that explains the extent to which the present oscillatory activity in a sender channel influences the future oscillatory activity in a specific receiver channel relative all possible receivers in the network. Using the proposed modeling approach, we have achieved some insights on learning in a rat engaged in a non-spatial memory task.

KEYWORDS : Electroencephalograms, Local field potentials, Brain effective connectivity, Multivariate time series, Neuroimaging, Vector Autoregressive model, Partial Directed Coherence.

¹Department of Statistics, University of California, Irvine, USA

²Department of Neurobiology and Behavior, University of California, Irvine, USA

³King Abdullah University of Science and Technology (KAUST), Saudi Arabia

⁴Corresponding author: hombao@uci.edu

1 Introduction

Connectivity between populations of neurons is a crucial component of understanding brain processes during cognition (e.g., memory and learning) and even during resting-state. Moreover, alterations in brain connectivity is widely believed to be implicated in a number of neurological and mental diseases such as obsessive compulsive disorder and Alzheimer’s disease. However, the underlying mechanisms of brain connectivity remain elusive. First, there is no set of universally-agreed measures for characterizing connectivity. Second, there are major statistical and computational challenges for modeling and analyzing multichannel brain signals – especially when the number of parameters is large which often happens when the number of spatial locations is large and/or the temporal lag for parametric models such as vector autoregressive (VAR) is large. Our contribution in this paper is a scalable approach to model connectivity for multichannel and potentially high dimensional brain physiological signals.

The work is motivated by our current collaborations with neuroscientists whose research requires developing a systematic statistical approach to study brain connectivity in rats while they were performing a memory task. The electrophysiological data recorded from rats are local field potentials (LFPs) and an example for one epoch is given in Figure 1. LFP signals have excellent temporal resolution (here we have 1000 observations per second). It is comparable to electroencephalograms (EEGs) in terms of temporal resolution and the nature of the signal (i.e., both capture electrical activity of the neurons). However, LFPs are recorded under invasive methods since these are obtained from electrodes that are chronically implanted inside the brain. Because LFPs are obtained from implanted electrodes, they have lower contamination compared to EEGs. They contain less non-neuronal physiological activity (e.g., muscular activity) and therefore possess a higher signal-to-physiological-noise ratio. One disadvantage of LFPs, however, is its limited utility in humans due to its invasive nature. However, these will continue to be a valuable tool for understanding brain function in animals which can then provide useful information for modeling brain function in humans. One of the challenges to fitting statistical models to LFPs is that the parameter space can be high dimensional. The number of recording leads in LFPs can range from 8-100; and the temporal order of parametric models such as vector autoregressive (VAR) models needs to be sufficiently large in order to accurately capture the dynamics in these complex processes.

In this paper we will develop a computationally scalable method for fitting high dimensional complex models that address two important goals in brain science: (1.) To identify the connectivity structure between channels in a

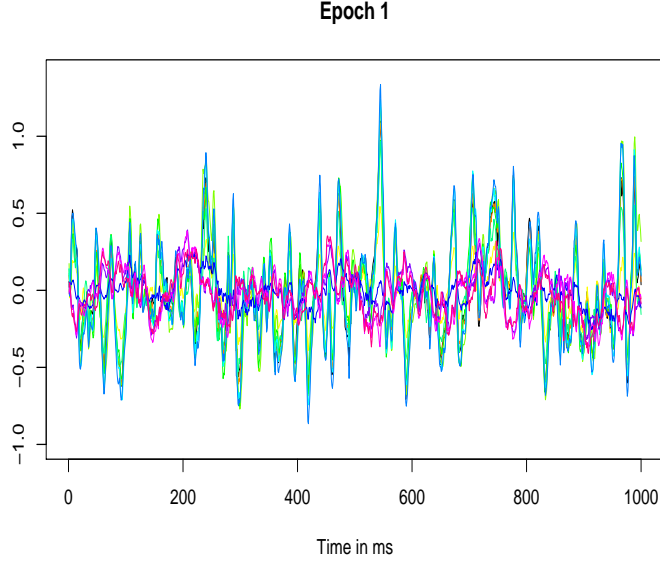


Figure 1: Local field potential (LFP) recording for one epoch; 1000 milliseconds; $T = 1024$.

brain network and (2.) To quantify both the strength and directionality of connectivity between these channels. Our approach is to fit a VAR model with potentially high temporal lag in order to more accurately capture complex lead-lag temporal dynamics between the channels or leads. Estimates of the VAR model will be obtained by a combination of regularization to maintain high specificity and least squares estimation to reduce bias and mean-squared error. The method will be applied to LFPs obtained from a rat in an olfaction study who was exposed to different odors.

To characterize connectivity in a multichannel LFP signal we shall use the vector autoregressive (VAR) model. A P -dimensional brain signal \mathbf{X}_t is said to follow a VAR model of order d , denoted $\text{VAR}(d)$, if it has the representation

$$\mathbf{X}_t = \Phi_1 \mathbf{X}_{t-1} + \dots + \Phi_d \mathbf{X}_{t-d} + \varepsilon_t \quad t = d+1, \dots, T \quad (1.1)$$

where $\Phi_k \in \mathcal{R}^{P \times P}$ are the autoregressive coefficient matrices and $\varepsilon_t \stackrel{iid}{\sim} N_P(\vec{0}, \Sigma)$. The dynamics of the process are determined by the autoregressive coefficient matrices $\{\Phi_k\}_{k=1}^d$ and spatial covariance matrix Σ . Thus, the VAR model provides a broad framework for capturing complex temporal and cross-sectional interrelationship among the time series (in particular, directionality of frequency-specific connectivity). Consequently it can be applied to model the Granger-causal relation between channels (Kamiński et al. [2001]).

To illustrate connectivity via the VAR matrix, consider Figure 2 and denote the LFP traces of brain region to be u -th and v -th channel. Then the entry Φ_ℓ^{uv} shows the impact of the input from v -th channel at time $t - \ell$ to brain

activity at u -th channel at the current time t . If $\Phi_\ell^{uv} = 0$ for all lags ℓ then, there is no connectivity between these two channels. A positive value indicates that the signal of v -th channel at time $t - \ell$, conditional on LFP values at other times, has positive linear dependence with u -th channel at time t . That is, an increase in activity in v -th channel leads to a increased future activity in u -th channel. Thus, the entries of $\{\Phi_k\}_{k=1}^d$ contain the information of brain connectivity between channels. In this paper, we shall use partial directed coherence (PDC) (Baccalá and Sameshima [2001], Baccalá and Sameshima [2014]) to characterize effective connectivity. This measure is more specific and provides more information than simply the coefficients of the VAR matrices. PDC is frequency-specific: it measures how an oscillatory activity (at a particular frequency band) at a present time in one channel may impact oscillatory activity of the same frequency band at another channel at a future time point.

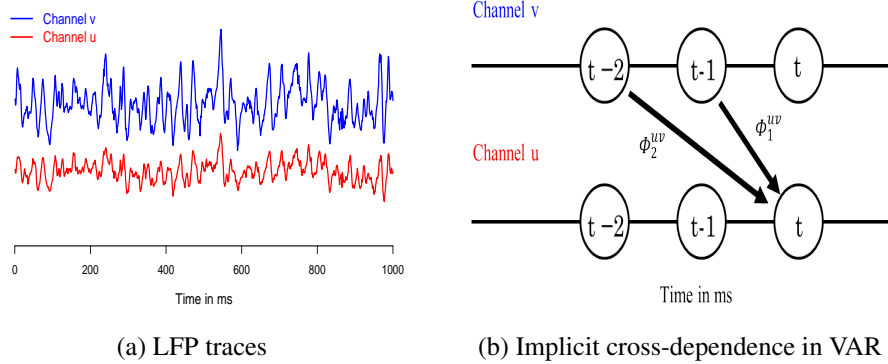


Figure 2: LFP traces and VAR

As noted above, effective connectivity between channels will be characterized by the VAR coefficient matrices. This is challenging because the parameter space of a VAR model for brain signals is usually high. For example, if we fit a VAR(10) model to 12 leads or channels, there are $10 \times 12^2 = 1440$ parameters in total to estimate, which subsequently require intensive computation. One could suggest fitting a model with low temporal lag in order to reduce the number of parameters. The problem with this suggestion, unfortunately, is that a low temporal order might miss potentially important features of the data such as multiple peaks in the spectra.

One classic estimation approach is via least squares which provides unbiased estimator for the elements of the VAR coefficient matrices but at the cost of high demand of computing. The least squares estimate (LSE) does not possess the specificity for coefficients with true value of zero. Hence it cannot provide an adequate answer to the first question about identifying effectively connected regions in brain network. Another common estimation approach is

the LASSO (least absolute shrinkage and selection operator) method which is a particular representative of the penalized regression family (Tibshirani [1996], Fu [1998], Zhao et al. [2009], Hesterberg et al. [2008]). Compared with LSE, the LASSO approach requires smaller computation time (Mairal and Yu [2012]). Most importantly, LASSO has higher specificity of zero-coefficients. The main limitation of the LASSO (and most regularization methods) is that the estimators of the non-zero coefficients are biased. Thus, it could lead to misleading results when investigating strength of brain effective connectivity. Inspired by the strengths of each of the two classical approaches (i.e., LSE and the LASSO), we propose to combine these in a two-step estimation procedure for the VAR coefficient matrices. Our method has inherited low bias for non-zero estimates and high specificity for zero-estimates from LSE and the LASSO separately. As a result, the proposed two-step method has higher specificity and significantly lower mean squared error (MSE) in the simulation study. At this stage, the full theoretical justification is being developed but the numerical experiments are encouraging.

A natural question to ask is whether or not the LASSO method is appropriate for fitting VAR models to brain signals. The answer lies in whether or not brain signals such as LFPs and EEGs indeed exhibit sparse connectivity structure. Due to volume conduction for EEGs, it is not likely that the connectivity structure between channels is sparse. However, LASSO aims to shrink many of the VAR coefficients to zero but this does not necessarily lead to a sparse connectivity structure. Keep in mind that a pair of channels are effectively disconnected only if *all* of the its corresponding VAR coefficients at *all* lags are estimated to be zero. Thus, imposing sparsity on the VAR coefficient matrices helps to weed out the less important parameters in the VAR model but does not oversimplify the connectivity structure.

The remainder of this paper is arranged as follows. In Section 2, we present the proposed hybrid LASSLS (LASSO+LSE) method followed by finite sample simulation studies in Section 3 and analysis of LFP sigals in Section 4 and the conclusion in Section 5.

2 A Proposed Two-Step LASSO-Least Squares Estimation Procedure for Fitting a VAR Model

First, we note that the VAR(d) model can be alternatively written in a form

$$\underbrace{\begin{bmatrix} (\mathbf{X}_T)' \\ \vdots \\ (\mathbf{X}_{d+1})' \end{bmatrix}}_Y = \underbrace{\begin{bmatrix} (\mathbf{X}_{T-1})' & \cdots & (\mathbf{X}_{T-d})' \\ \vdots & \ddots & \vdots \\ (\mathbf{X}_d)' & \cdots & (\mathbf{X}_1)' \end{bmatrix}}_{\mathbb{X}} \underbrace{\begin{bmatrix} (\Phi_1)' \\ \vdots \\ (\Phi_d)' \end{bmatrix}}_B + \underbrace{\begin{bmatrix} (\varepsilon_T)' \\ \vdots \\ (\varepsilon_{d+1})' \end{bmatrix}}_E. \quad (2.1)$$

Next, denote $Y = [y_1, y_2, \dots, y_P]$, $B = [b_1, b_2, \dots, b_P]$, $E = [e_1, e_2, \dots, e_P]$. Denote the k^{th} column vector of the matrices Y , B and E ($k = 1, 2, \dots, P$) to be y_k , b_k , e_k . Then we have

$$\underbrace{y_k}_{m \times 1} = \underbrace{\mathbb{X}}_{q \times 1} \underbrace{b_k}_{m \times 1} + \underbrace{e_k}_{m \times 1}, \quad e_k \stackrel{indep}{\sim} N_m(\vec{0}, \sigma_{kk} I_m) \quad (2.2)$$

where $m = T - d$, $q = P \times d$, and σ_{kk} is the k^{th} diagonal element of Σ which is the covariance matrix of e_k .

Note that Equation (2.1) is finally decomposed into many sub-linear regression problems of estimating $\{b_k\}_{k=1}^P$ in a parallel manner and all the entries of connectivity matrices are included by $\{b_k\}_{k=1}^P$.

2.1 Least Squares Estimation (LSE)

To fit a linear regression model, the most common approach is via least squares estimation so that the least squares estimator \hat{b}_k satisfies

$$\hat{b}_k = \underset{b_k \in \mathbb{R}^q}{\operatorname{argmin}} \|y_k - \mathbb{X}b_k\|^2 \quad (2.3)$$

which gives the unbiased estimator $\hat{b}_k = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'y_k$. Some papers Han and Liu [2013] argue that in high dimensional case the number of parameters q can be larger than the number of observations m , thus this method has limitations due to the nonsingular matrix $\mathbb{X}'\mathbb{X}$. However, at this point, we do not worry about when analyzing the LFP data since normally we have replicated measurements from multiple epochs. The biggest problem here is that LSE has poor specificity for coefficients with true value of zero. It always produces estimates that are very close to zero rather than exactly zero, which reflects non-connectivity between channels. Indeed when EEG channels are not effectively connected with each other, then an excess non-zero estimate could lead to incorrect characterizations of connectivity through partial directed coherence (PDC). Moreover, even a trivial amount of bias for one coefficient, when added across thousands of coefficients, can produce large mean squared error (as demonstrated in the simulation study).

2.2 LASSO family estimation

In order to overcome the problem of non-specificity by the LSE method, recent attention has been focused on the family of penalized regression models as viable solutions to this problem. One of the well known methods of this family is LASSO regression (with L_1 penalty term). The estimates given by LASSO are the solution the minimization of the criterion

$$\tilde{b}_k = \underset{b_k \in \mathbb{R}^q}{\operatorname{argmin}} \|y_k - \mathbb{X}b_k\|^2 + \lambda \|b_k\|_1 \quad (2.4)$$

The penalty term will force a lot of excess non-zero estimates to exact zero, which provides good estimate for the sparsity of the VAR coefficient $\{\Phi_k\}_{k=1}^d$ which could greatly simplify the calculation of connectivity measures (e.g., PDC) by focusing only on the the more important coefficients. In the implementation of the algorithms for LASSO, we took advantage of the results demonstrated by Friedman et al. [2010] where estimation of generalized linear models with convex penalties can be handled by cyclical coordinate descent and computed along a regularization path. The price of LASSO is that the non-zero estimates are biased of true values which leads to incorrect estimates of the strength of connectivity between channels (PDC).

2.3 LASSLE: Proposed Two-step estimation method

Motivated by both the advantages and limitation of each of the previous approaches, we propose a two-step procedure to estimate VAR model. Our method consists of these two steps:

Step 1. Apply LASSO to identify entries in $\{\Phi_k\}_{k=1}^d$ whose estimates are not set to 0.

$$\widehat{S}_k = \{j \in \{1, \dots, q\} : \widehat{b}_k^j \neq 0\} \quad (2.5)$$

Step 2. Fit LSE with the constraint that “zero” entries estimates from Step 1 are fixed to 0

$$\tilde{b}_{kLAS} = \underset{b_k : b_k^j = 0, j \in \widehat{S}_k^c}{\operatorname{argmin}} \|y_k - \mathbb{X}b_k\|_2^2 \quad (2.6)$$

To obtain the optimal tuning parameter λ , we employ a K -fold cross-validation test in Step 1. A sequence of candidates of $\{\lambda_k\}$ will be pre-specified and the optimal value is selected such that the average of prediction error on test data is minimized.

LASSLE Algorithm

- 1: **procedure** TWO-STEP ESTIMATION
 - 2: *Step 1:*
 - 3: Generate a sequence of (d, λ) and randomly divide data to K folds
 - 4: For a possible choice of (d, λ) , leave one fold as test data at each time
 - 5: Train 2.2 with LASSO method on other folds and compute $\{\widehat{\Phi}_k\}_{k=1}^d$ for $\{\Phi_k\}_{k=1}^d$
 - 6: Based on $\{\widehat{\Phi}_k\}_{k=1}^d$, calculate prediction error on test set and finally take average
 - 7: Select (d, λ) with the lowest average prediction error
 - 8: Obtain estimate $\{\widehat{b}_k\}_{k=1}^P$ for $\{b_k\}_{k=1}^P$ in 2.2 of lag d using LASSO method with λ
 - 9: *Step 2:*
 - 10: **if** $\widehat{b}_k^j = 0$ **then**
 - 11: Set $b_k^j = 0$.
 - 12: **if** $\widehat{b}_k^j \neq 0$ **then**
 - 13: Keep b_k^j .
 - 14: Obtain estimate $\{\widetilde{b}_k\}_{k=1}^P$ for $\{b_k\}_{k=1}^P$ in 2.2 with LSE under above constriction
 - 15: Obtain estimate $\{\widetilde{\Phi}_k\}_{k=1}^d$ for $\{\Phi_k\}_{k=1}^d$ by arranging $\{\widehat{b}_k\}_{k=1}^P$
-

2.4 Theoretical Consideration

For linear regression, under Irrepresentable Condition¹, $\{\widehat{b}_k\}_{k=1}^P$ have sign consistency assured by LASSO estimator

(Zhao and Yu [2006]), which means for sufficient large sample size $T - d$

$$\mathbb{P}(\text{sgn}(\widehat{b}_k) = \text{sgn}(b_k)) \rightarrow 1 \quad (2.7)$$

where $\text{sgn}(b_k)$ is the sign function with value of 1, 0 or -1 corresponding to $b_k > 0$, $b_k = 0$ or $b_k < 0$ respectively.

Therefore, $\mathbf{P}(\widehat{S}_k \neq S_k) \rightarrow 0$, which implies high specificity of true truly zero VAR coefficients. Then our inaccurate non-zero estimate will be polished in Step 2. Since we put a constraint for LSE in Step 2, the computing is much simplified compared with merely LSE. Moreover, the bias and mean squared error of LASSLE estimator will be bounded (Liu et al. [2013])

$$\|E(\widetilde{b}_{kLAS}) - b_k\|_2^2 \leq 2\mathbf{P}(\widehat{S}_k \neq S_k)\{O(\frac{1}{m}) + \|b_k\|_2^2 + \tau\sigma_{kk}\} \quad (2.8)$$

$$E\|\widetilde{b}_{kLAS} - b_k\|_2^2 \leq 2\frac{\sigma_{kk}}{m}\text{tr}(\Psi_{11}^{-1}) + \sqrt{\mathbf{P}(\widehat{S}_k \neq S_k)\{O(\frac{1}{m}) + \|b_k\|_2^2 + \tau\sigma_{kk}\}} \quad (2.9)$$

Thus, our non-zero estimates are almost unbiased, which is significantly improved from LASSO. Final estimates

given by LASSLE in simulation study have substantially lower general mean squared error. Thus our approach is

¹Assume $b_k = (b_k^1, \dots, b_k^J, b_k^{J+1}, \dots, b_k^q)^T$, where $b_k^j \neq 0$ for $j = 1, \dots, J$ and $b_k^j = 0$ for $j = J+1, \dots, q$. Let $b_k^{(1)} = (b_k^1, \dots, b_k^J)^T$ and $b_k^{(2)} = (b_k^{J+1}, \dots, b_k^q)^T$. Denote Gram matrix $\Psi = \frac{1}{n}\mathbb{X}'\mathbb{X} = \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix}$, then Irrepresentable Condition is satisfied if there exists a positive constant vector η , such that $|\Psi_{21}(\Psi_{11})^{-1}\text{sgn}(b_k^{(1)})| \leq 1 - \eta$.

able to both indicate the most important effective connectivity and give a more accurate and precise estimate of the strength of connectivity.

2.5 Measure of Dependence

In this section, we enumerate the different measures of dependence between components of a multivariate time series (or between different channels) using the VAR model. First, a P -channel time series, denoted $\{\mathbf{X}_t = (\mathbf{X}_t^1, \dots, \mathbf{X}_t^P)'\}$, $t = 1, 2, \dots$, is weakly stationary if

(a.) $E(\mathbf{X}_t)$ is constant over all time t , and

(b.) the autocovariance function matrix

$$\Gamma(h) = \begin{pmatrix} \gamma_{11}(h) & \gamma_{12}(h) & \dots & \gamma_{1P}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) & \dots & \gamma_{2P}(h) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{P1}(h) & \gamma_{P2}(h) & \dots & \gamma_{PP}(h) \end{pmatrix}$$

depends only on the lag h , where $\gamma_{uv}(h) = \text{cov}(X_t^u, X_{t+h}^v)$ for all pairs of channels $u, v = 1, \dots, P$.

Moreover, if the sequence of auto- and cross-covariance between any pair of channels u and v is absolutely summable,

i.e., $\sum_{h=-\infty}^{\infty} |\gamma_{uv}(h)| < \infty$, then we define the spectral density matrix of $\{\mathbf{X}_t\}$ as

$$f(\omega) = \sum_{h=-\infty}^{\infty} \Gamma(h) e^{-2\pi i \omega h} - 1/2 \leq \omega \leq 1/2 \quad (2.10)$$

This matrix is $P \times P$ whose diagonal elements $f_{uu}(\omega)$ are the auto-spectrum of the channels at frequency ω and the off-diagonal elements $f_{uv}(\omega)$ are the cross-spectrum of channels u and v at frequency ω .

The first dependency measure that we will consider is coherency. Coherency between the u -th and v -th channels at frequency ω , is defined as

$$\rho_{uv}(\omega) = \frac{f_{uv}(\omega)}{\sqrt{f_{uu}(\omega)} \sqrt{f_{vv}(\omega)}}. \quad (2.11)$$

One can interpret coherency as the cross-correlation between the ω -oscillatory component in channel u and the ω -oscillatory component in channel v (Ombao and Van Bellegem [2008]).

The second dependence measure is coherence. Coherence between the u -th and v -th channels at frequency ω , is defined as

$$\rho_{uv}^2(\omega) = \frac{|f_{uv}(\omega)|^2}{f_{uu}(\omega) f_{vv}(\omega)}. \quad (2.12)$$

When $\rho_{uv}^2(\omega)$ is close to 1 then both channels u and v share a common ω -oscillatory activity. Moreover, when the cross-correlation between the u and v channels is 0 at all time lags, then the coherency (and coherence) between these

channels at *all* frequencies is 0. A large coherence value between channels u and v could be due to direct connectivity between these two channels or could be due to the intervening effect of other channel(s). To measure the strength of connectivity between a pair of channels – with the effect of all intervening channels removed – we shall use partial coherence.

The third dependency measure is partial coherence. Define the matrix $g(\omega) = f^{-1}(\omega)$ and denote the diagonal elements as $g_{pp}(\omega)$. Let $h(\omega)$ be a diagonal matrix whose elements are $g_{pp}^{-1/2}(\omega)$. Define the matrix $C(\omega)$ to be

$$C(\omega) = -g(\omega)h(\omega)g(\omega) \quad (2.13)$$

Then, the partial coherence between the u -th and v -th channels is the modulus squared of the (u, v) -th element of $C(\omega)$ (Fiecas et al. [2010], Fiecas and Ombao [2011])

$$\tilde{\rho}_{uv}^2(\omega) = |C_{uv}(\omega)|^2 \quad (2.14)$$

We now present the fourth dependency measure which is partial directed coherence (Baccalá and Sameshima [2001], Baccalá and Sameshima [2014]). Consider a VAR(d) model given by Equation 1.1, define

$$A(\omega) = I - \sum_{k=1}^d \Phi_k \exp(-i2\pi\omega k/\Omega) \quad (2.15)$$

be the Fourier transform of $\{\Phi_k\}_{k=1}^d$ at frequency ω , where Ω is the sampling frequency. The partial directed coherence (PDC) between the u -th and v -th channels at frequency ω is defined as

$$\pi_{uv}^2(\omega) = \frac{|A_{uv}(\omega)|^2}{\sum_{m=1}^P A_{mv}(\omega)(A_{mv}(\omega))^*} \quad (2.16)$$

which measures the direct influence from channel v to channel u conditional on all the outflow from channel v . PDC gives an indication on the extent to which present frequency-specific oscillatory activity from a sender channel explains future oscillatory activity in a specific receiver channel relative to all channels in the network.

2.6 Model Selection

To determine the best order \hat{d} of VAR, we first use least square estimation to obtain $\{\widehat{\Phi}_k\}_{k=1}^{d_j}$ for candidate order $\{d_j\}_{j=1}^J$. We search among a class of reasonable temporal lag orders. From our analysis of LFPs where there are usually less than 4 peaks in the spectrum, it would be reasonable to use an upper bound of 12 as the temporal lag order. Then we calculate the sum of squared errors

$$SSE = \sum_{t=d_j+1}^T (\mathbf{X}_t - \sum_{k=1}^{d_j} \widehat{\Phi}_k \mathbf{X}_{t-k}) (\mathbf{X}_t - \sum_{k=1}^{d_j} \widehat{\Phi}_k \mathbf{X}_{t-k})' \quad (2.17)$$

Consequently the conditional MLE of the error covariance matrix Σ is

$$\hat{\Sigma} = SSE/(T - d_j) \quad (2.18)$$

which is analogous to univariate regression case. There are three information criteria to evaluate d_j are, respectively, the Akaike Information criterion (AIC), Bayesian Information criterion (BIC) and Hannan-Quinn information criterion (HQC)

$$AIC(d_j) = \log |\hat{\Sigma}| + 2/T(p^2 d_j) \quad (2.19)$$

$$BIC(d_j) = \log |\hat{\Sigma}| + \log T/T(p^2 d_j) \quad (2.20)$$

$$HQC(d_j) = \log |\hat{\Sigma}| + 2 \log \log T/T(p^2 d_j) \quad (2.21)$$

The optimal order for each criterion, denoted \hat{d} is the minimizer of the cost functions and thus gives the optimal balance between fit (as measured by SSE) and model complexity (as expressed by the penalty terms). It has been noted that $\hat{d}^{BIC} \leq \hat{d}^{AIC}$ when $T \geq 8$, $\hat{d}^{BIC} \leq \hat{d}^{HQC}$ for all T , and $\hat{d}^{HQC} \leq \hat{d}^{AIC}$ when $T \geq 16$ (Ivanov et al. [2005]).

3 Simulation Study

3.1 Simulation design

To compare the performance of the proposed LASSLE approach with the classical methods (i.e., LSE and LASSO), we conducted a simulation study of VAR(d) model for two different brain network types - both high dimensional (i.e., many nodes). We call the first type “Cluster”, of which auto-connectivity of EEG channels within same brain region distinguish each region as different small networks. The “Cluster” type of network has high level local and global connectivity efficiency. In Figure 3, channels (red points) are located in four brain regions, while the gray line between two red points indicates connectivity at channel level. Auto-connectivity inside each region makes channels from the same region connect like a cluster, and cross-connectivity between brain regions determines whether these clusters are connected with each other. For example, region 2 is independent from other regions, but region 1 and region 4 are connected due to the cross-connectivity at region level.

In the second type “Scale-free” shown in Figure 4, there is no significant auto-connectivity or cross-connectivity at the region level, but all the EEG channels are connected within the network. Each channel has several connections with other channels on average, with the exception that a few channels are heavily connected. The internal philosophy

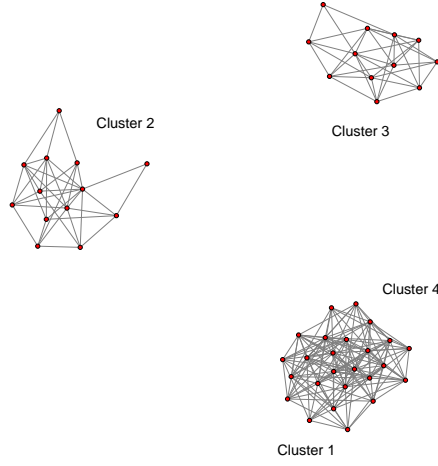


Figure 3: 50 EEG channels of “Cluster” type

is that these channels play a central role in the organization of entire brain network, as they are mostly responsible for the connectivity efficiency.

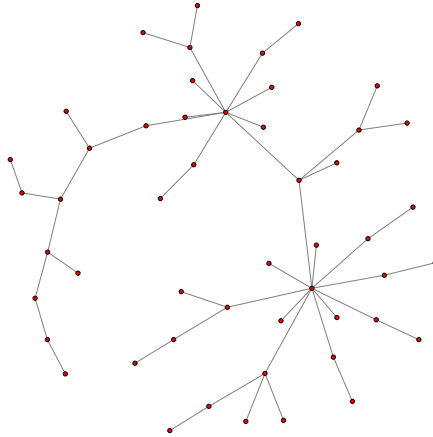


Figure 4: 50 EEG channels of “Scale-free” type

For both network types, we use 1.1 to generate time series data sets. Φ_1 of setting $\{P = 50, d = 1\}$ is visualized in Figure 5 and Figure 6. Each small rectangle represents the non-zero entry of Φ_1 and different colors indicate

different values according to the color bar. The blank parts of connectivity matrix are the zero entries. In addition, ε_t follows a Gaussian distribution and the covariance matrix is non-diagonal. We run $N = 1000$ simulations for each VAR setting respectively and the time series data for each channel contains $T = 10000$ time points. Then we apply LSE, LASSO and our LASSLE method to estimate coefficient matrices, and compare their results with two important criteria. The first one is how successful the estimate identifies entries with true value of zero, which is shown by the visualization of absolute difference between true coefficient matrices and estimated ones. The second criteria is their mean squared error, defined as

$$MSE = \frac{\sum_{n,k,i,j} (\Phi_k^{ij} - \widehat{\Phi}_k^{ij})^2}{N} \quad (3.1)$$

where $\{\Phi_k^{ij}\}_{k=1}^d, \{\widehat{\Phi}_k^{ij}\}_{k=1}^d$ represent entries of $\{\Phi_k\}_{k=1}^d$ and $\{\widehat{\Phi}_k\}_{k=1}^d$ respectively. Lower MSE indicates better centering at true connectivity matrix.

3.2 Simulation results

Due to the display limit of high dimensional matrix, we only demonstrate visualized results of setting $\{P = 50, d = 1\}$.

3.2.1 Results of “Cluster”

In this setting, 50 channels represent measurements in four brain regions and only region 1 and region 4 have cross-connectivity. In connectivity matrix, all diagonal entries are set 0.5, and other non-zero entries are randomly assigned either 0.1 or -0.1 (shown in Figure 5(a)). Figure 5(a),(b),(c) yields the absolute difference between each estimate and true connectivity matrix. The color of small rectangles ranging from white to red indicates the value of absolute difference of each entry. The blank part of the matrix implies that the estimate has given correct zero-estimate for those true zero entries so that there is no need to distinguish the difference with color. Table 1 demonstrates the MSE results of all three methods.

Table 1: Comparison of MSE for “Cluster” data

VAR Setting			MSE $\times 10^{-3}$		
Parameters	Dimension	Lag	LSE	LASSO	LASSLE
100	10	1	8	93	<u>4</u>
500	10	5	58	326	<u>36</u>
1,000	10	10	134	412	<u>82</u>
2,500	50	1	176	464	<u>24</u>
5,000	50	2	457	739	<u>93</u>
10,000	100	1	697	1016	<u>65</u>

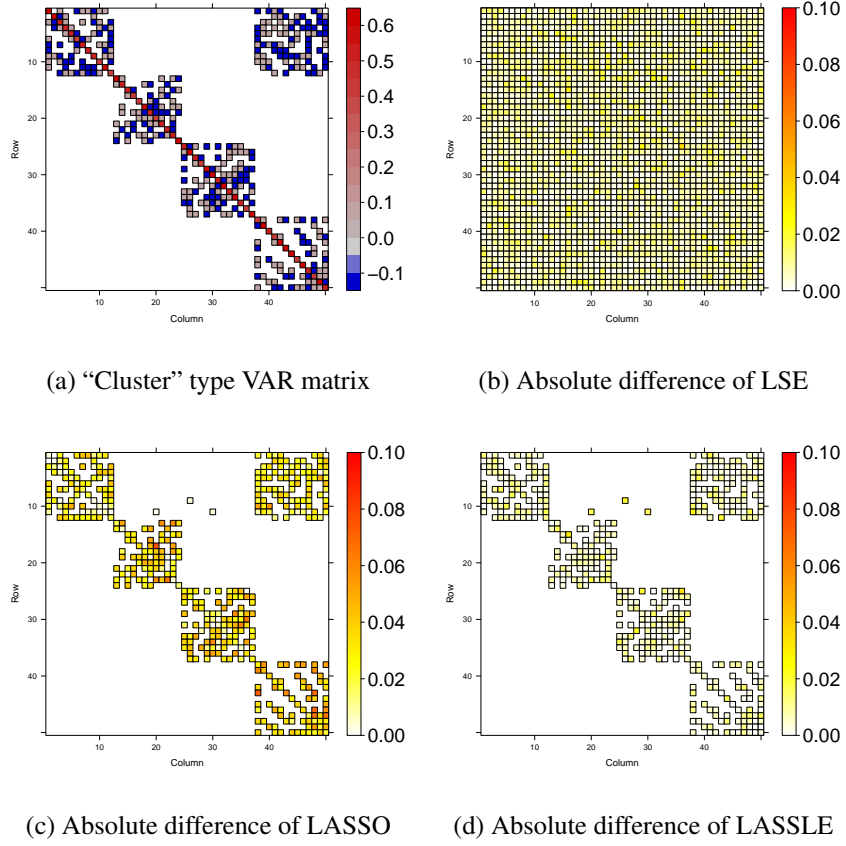


Figure 5: Absolute difference of different methods

3.2.2 Results of “Scale-free”

To generate “Scale-free” type data, we assign 0.5 to all diagonal entries of connectivity matrix, and 0.1 or -0.1 randomly to other non-diagonal entries with small probability (seen in Figure 6(a)). Figure 6(b),(c),(d) shows the visualized estimate results given by LSE, LASSO and our approach separately. MSE comparison can be found in Table 2.

Table 2: Comparison of MSE for “Scale-free” data

VAR Setting			MSE $\times 10^{-3}$		
Parameters	Dimension	Lag	LSE	LASSO	LASSLE
100	10	1	7	86	<u>2</u>
500	10	5	53	158	<u>16</u>
1,000	10	10	130	472	<u>72</u>
2,500	50	1	185	439	<u>11</u>
5,000	50	2	480	877	<u>37</u>
10,000	100	1	762	921	<u>19</u>

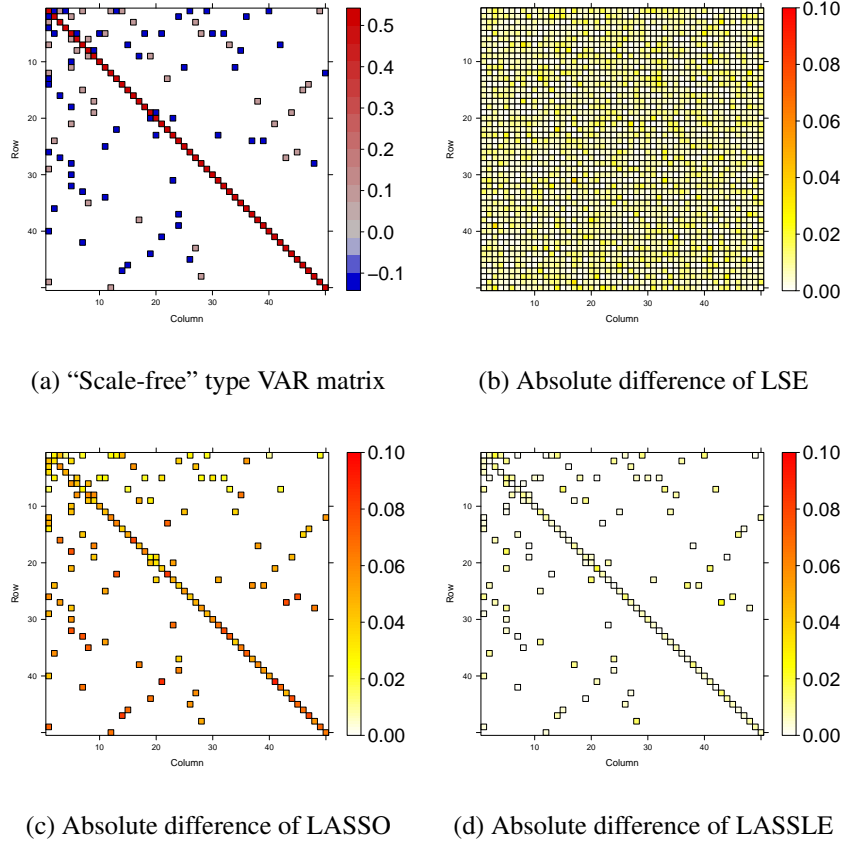


Figure 6: Absolute difference of different methods

3.3 Discussion

From visualized results, we can find that LSE is unable to give specificity for true zero coefficients, since its estimates do not contain blank rectangles. However, its estimate has general lower bias across all the entries, which is implied by the light color of absolute difference. LASSO is able to identify most true zero entries, but darker color of rectangles indicate that this method has high bias for the estimate. Our method, constrained with LASSO in Step 1, has inherited the specificity of true zero values from LASSO, consequently can capture true zero values as well as LASSO. Thus in the sense of specificity of true zero, the comparison result is: $\text{LASSLE} = \text{LASSO} \gg \text{LSE}$.

For another important criteria MSE, the proposed LASSLS approach has substantially advantage over LSE and LASSO. In 50-channel “Cluster” settlement, the MSE given by LSE and LASSO are approximately 5 times and 10 times the MSE provided by our method. Also, in 50-channel “Scale-free” settlement, LSE and LASSO provide at least 10 times and 20 times higher MSE compared with LASSLE. With the increase of dimensions, the advantage

is also increasing geometrically. Therefore, LASSLE performs better with respect to the MSE criterion compared to both LSE and LASSO.

4 Application to Effective Connectivity in Multichannel LFPs

In this section, we will fit a VAR model to rat LFPs recorded in a nonspatial sequence memory task and apply the LASSLE method to estimate the VAR parameters and consequently partial directed coherence. Our objective is to study the alternation of their neural activity in hippocampal region CA1, which is clinically meaningful as sequence memory shows strong behavioral parallels in rats and humans (Allen et al. [2014]) and depends on the hippocampus for both species (Quirk et al. [2013]).

4.1 Data Description

In the experiment, rats were presented with repeated sequences of five odors in a single odor port. They were trained to identify whether each odor was presented "in sequence" (by holding their nose poke) or "out of sequence" (by withdrawing their nose poke) to receive a water reward. Well trained rats continued to be tested while LFPs in their CA1 pyramidal layer of the dorsal hippocampus were recorded (Allen et al. [2016]).

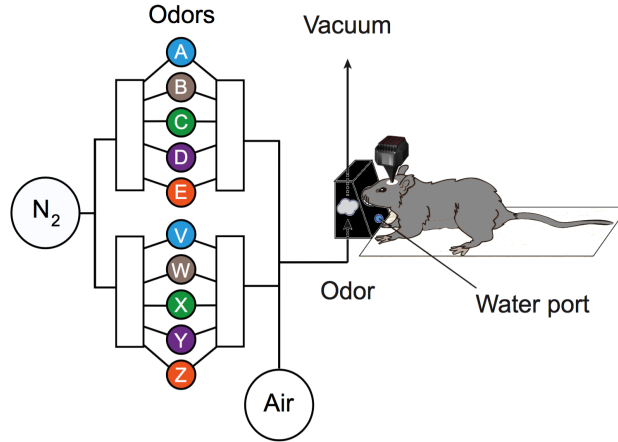


Figure 7: A nonspatial sequence memory experiment of rats. Rats were presented with repeated sequences of five odors (A,B,C,D and E) in a single odor port. Each odor presentation was initiated by a nosepoke and rats were required to correctly identify the odor as either InSeq (ABCDE) by holding their nosepoke or OutSeq (e.g.,ABDDE) by withdrawing their nosepoke.

The original LFPs consists of 23 tetrodes (location shown in Figure 8) and $n = 247$ epochs, where $n = 219$ epochs are "in sequence" (InSeq) and $n = 28$ epochs are "out of sequence" (OutSeq). Each epoch is recorded

roughly 1 second with sampling frequency of 1000 Hz and thus has $T = 1000$ time points. We focus on analyzing LFPs from $P = 12$ tetrodes, of which the corresponding with original tetrodes is demonstrated in Table 3.

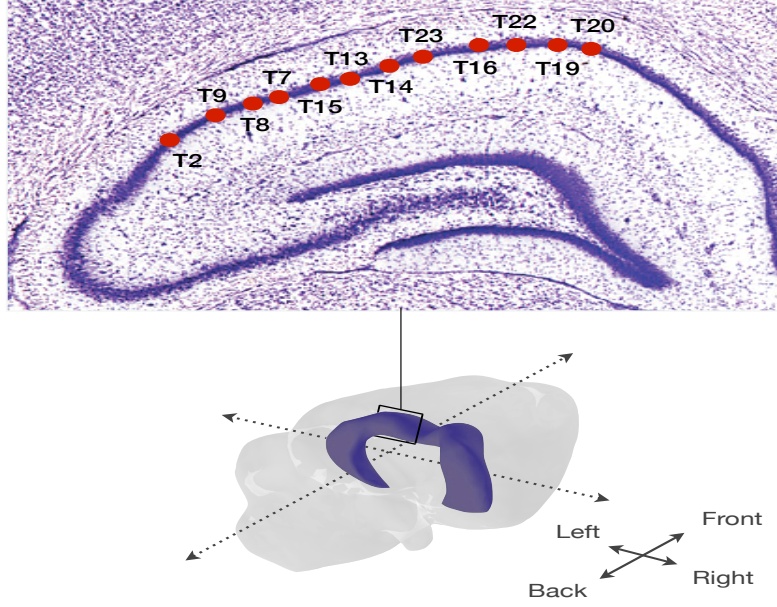


Figure 8: Location of tetrodes placed in the hippocampus of rats.

Table 3: Mapping between tetrodes in this study and tetrodes in the experiment

Studied tetrodes	Original tetrodes
Tetrode 1	T13
Tetrode 2	T14
Tetrode 3	T15
Tetrode 4	T16
Tetrode 5	T19
Tetrode 6	T20
Tetrode 7	T22
Tetrode 8	T23
Tetrode 9	T2
Tetrode 10	T7
Tetrode 11	T8
Tetrode 12	T9

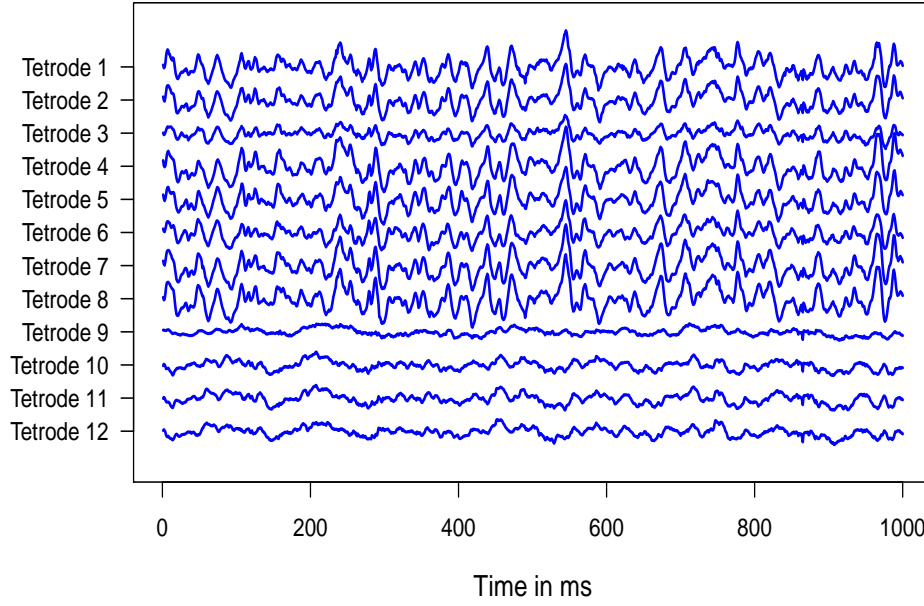


Figure 9: LFPs of epoch 1 in this study. Time series of LFPs from tetrode 1 to tetrode 8 display highly similar temporal pattern, while time series of the rest tetrodes are highly similar. This is because tetrodes near each other are likely to behave more similarly than those that are far apart.

Figure 10 (left) displays the LFPs from tetrode 7 during the first 30 epochs; the boxplots of its auto-correlation function (ACF) across all 247 epochs; and the boxplots of the partial auto-correlation function (PACF) across all epochs. We observe that the boxplots of ACF fail to drop to zero even after very long lags and there is a cyclic behavior in the pattern. Both of these could be evidence of non-stationarity (or long-memory). These suggest pre-processing the data by taking a first order difference. The results of LFPs after differencing are shown on the right of Figure 10. Compared to previous plots, the ACF boxplots eventually drop to zero with smaller IQR, which means that the processed data look more stationary and that the correlation drops to zero faster than the original LFP. Therefore we will fit the VAR model to the first order differenced LFPs.

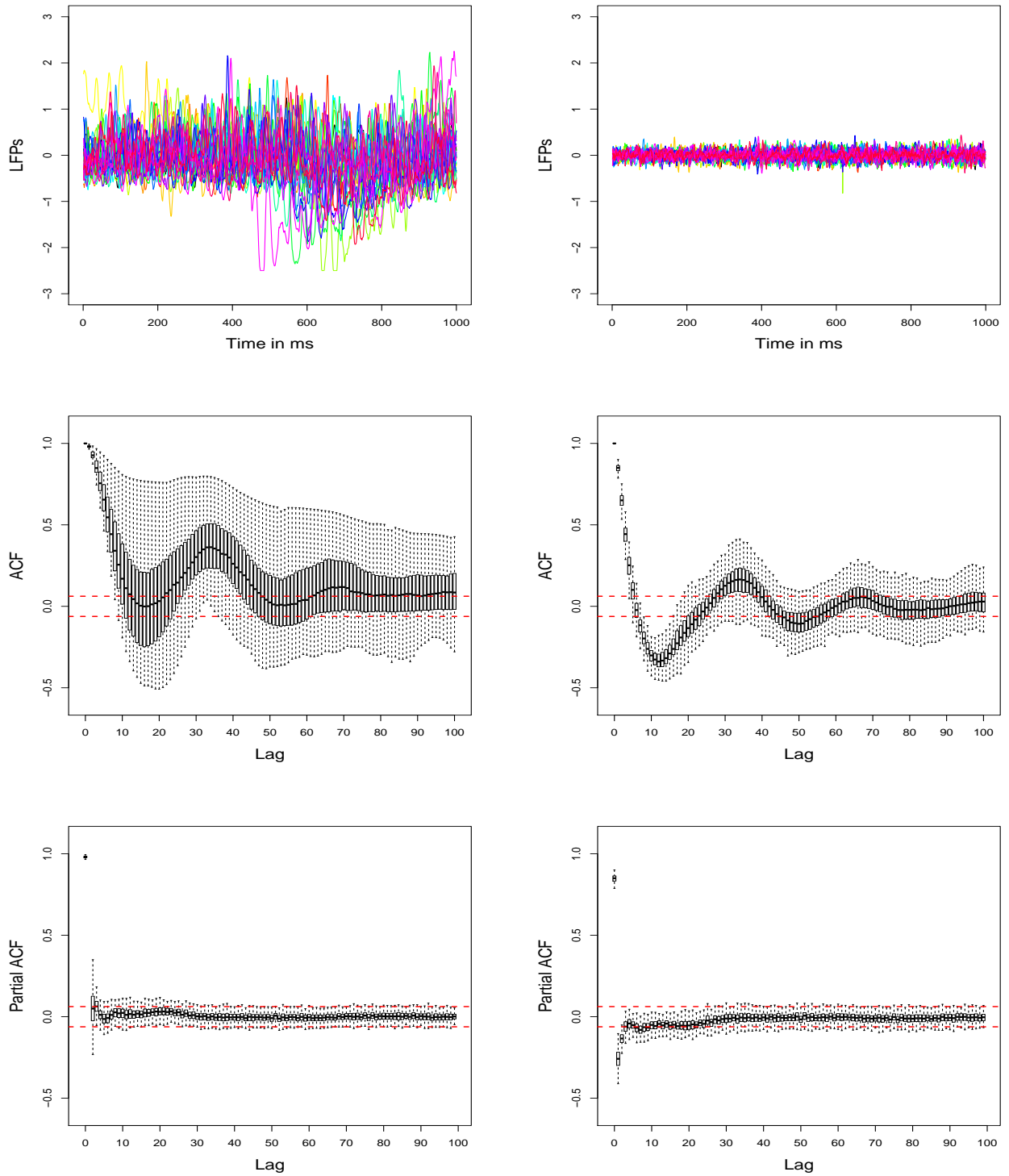


Figure 10: Top: The LFPs time series plots of the first 30 epochs at tetrode 7 before (left) and after (right) processing. Middle: The boxplots of auto-correlation function (ACF) from tetrode 7 before (left) and after (right) processing across all epochs. Bottom: The boxplots of partial auto-correlation function (PACF) from tetrode 7 before (left) and after (right) processing across all epochs.

4.2 Preliminary analysis of epoch 1

We first demonstrate fitting the VAR model to epoch 1. To select the best lag order \hat{d} , we first fit a $\text{VAR}(d_j)$ model to candidate orders $d_j \in \{1, 2, \dots, 12\}$ and use LSE to estimate the coefficient matrices. Then we apply 2.17, 2.18, and 2.19 to compute AIC for each candidate order d_j . The best order (or the minimizer of AIC) was $\hat{d} = 3$. Consequently we have 3 connectivity matrices of 12 dimensions to estimate. After estimating $\{\Phi\}_{k=1}^{\hat{d}}$ by LASSLE method, we use matrix-imaging to demonstrate our estimation results, in which small squares represent the entries of connectivity matrix. Blanks are assigned to entries whose value is zero, so non-dependence between tetrodes is easy to tell. For entries whose value is not zero, we assign them colors with red for positive value and blue for negative value, and the strongness of dependence is implied by the color-key. Figure 11 shows the estimates of Φ_1 , Φ_2 and Φ_3 for epoch 1. We can see that most diagonal entries of Φ_1 are either red or orange, which implies that signals have strong positive auto-dependence. In addition, entries in column 9 of Φ_1 are mostly blue, which could be an evidence that signal of tetrode 9 at time $t - 1$ has significant negative dependence with signals from other tetrodes at time t . Compared to Φ_1 , more than half entries in Φ_2 are blank, suggesting there is no auto- and cross-dependence between those tetrodes at time $t - 2$ and at time t . Column 4 and 6 in Φ_2 are light blue, which implies weak negative dependence between signals from tetrode 4, 6 at time $t - 2$ and signals from other tetrodes at time t . Also, we believe there is strong positive dependence between signals from tetrode 11 at time $t - 2$ and signals from tetrode 1,2,4,5,6,7 at time t as the color of column 11 in Φ_2 is orange. However, most entries of Φ_3 are blank and limited non-zero estimate are close to zero, which implies that the dependence between LFPs at time $t - 3$ and time t is almost none.

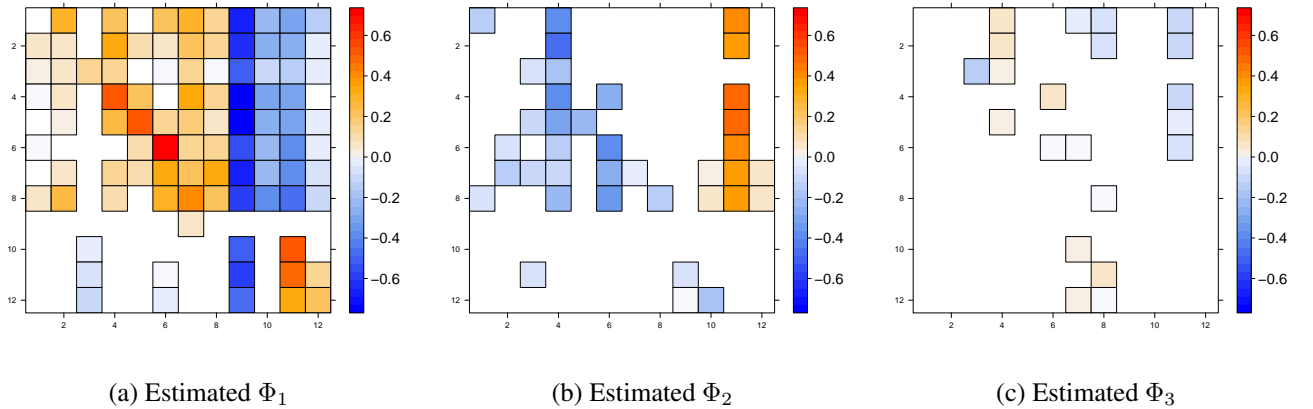


Figure 11: Estimated connectivity matrices of epoch 1

Next we use Equations 2.15 and 2.16 to calculate partial directed coherence in epoch 1. Particularly, we focus on PDC at the δ band (0-4 Hertz), θ band (4-8 Hertz), α band (8-12 Hertz), β band (12-32 Hertz) and γ band (32-50 Hertz) in current study. To estimate PDC at specific frequency *band*, we calculate the average of estimates of PDC over all singleton frequencies in that band. Figure 12 demonstrates the estimated PDC results of these frequency bands in epoch 1. Since there is only slight change on the estimated PDC across different frequency bands, we use the results of γ band (shown in Table 4) as representative to explain the PDC of epoch 1. For tetraode 1, 2, 3, 4, 5, 10 and 12, over 80% of their information flow back to themselves while most of their information flow to other tetrodes are very close to 0. More specifically, tetraode 2 has 9.3% information that flow to tetraode 1, and 11.1% information of tetraode 5 flows to tetraode 4. This implies that they tend to have communication with certain tetrodes instead of the entirety. Unlike these tetrodes, tetraode 6, 7, 8, 9 and 11 has significant amount of information flowing to other tetrodes. For example, the proportion of current tetraode 9 that is explained by its own past is only about 19.3%. This could be an evidence that these tetrodes play an important role of passing information to other tetrodes while the rat was engaged in a non-spatial memory task. Estimated PDC from tetraode 1-8 (sender) to tetraode 9-12 (receiver) are almost none (the blank on the bottom left of PDC), which suggest that previous oscillatory activity at γ frequency of tetraode 1-8 can hardly explain future oscillatory activity at γ frequency of tetraode 9-12 as they are far apart in spatial distance.

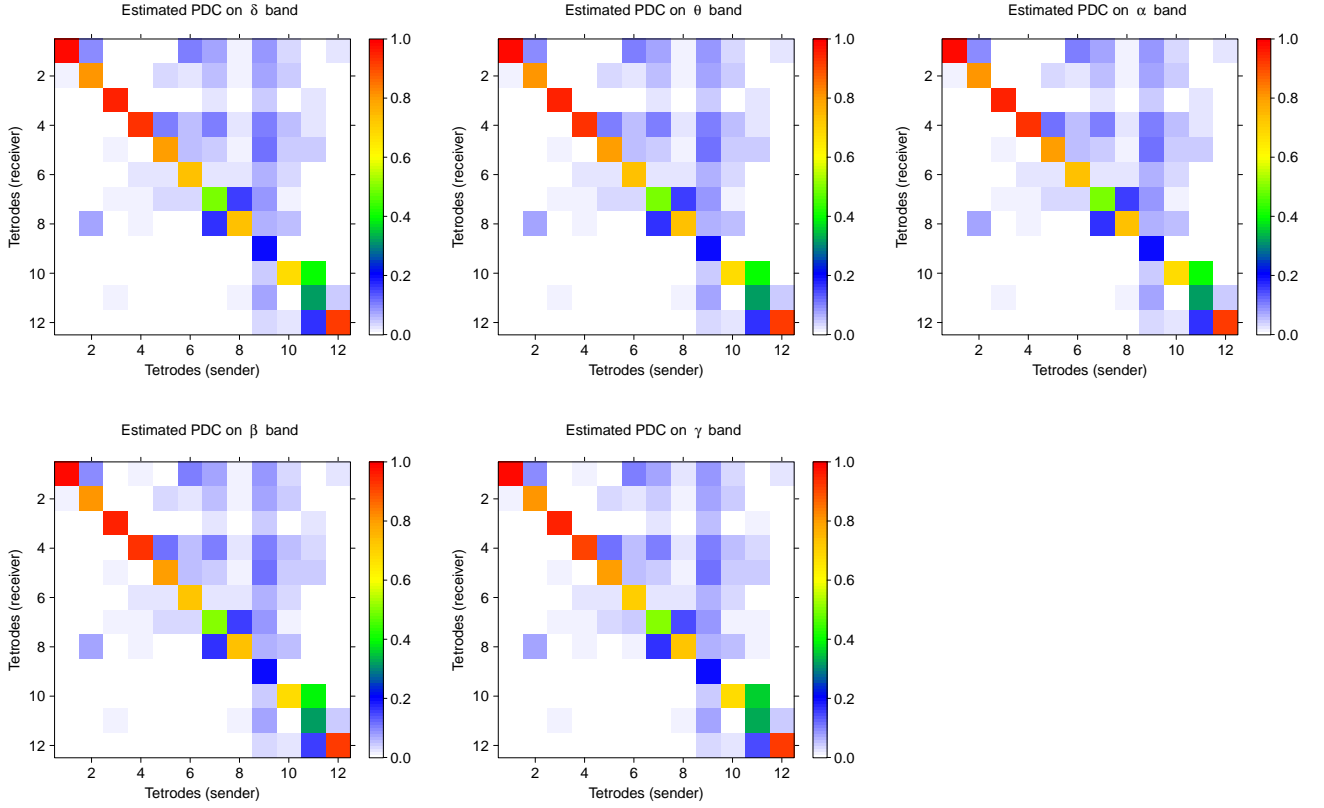


Figure 12: Estimated PDC of epoch 1

Tetrode	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
#1	0.983	0.093	0.000	0.008	0.000	0.106	0.074	0.018	0.082	0.039	0.002	0.025
#2	0.010	0.809	0.000	0.000	0.030	0.019	0.050	0.010	0.073	0.046	0.005	0.001
#3	0.003	0.008	0.955	0.003	0.000	0.000	0.023	0.000	0.050	0.008	0.020	0.002
#4	0.000	0.008	0.000	0.938	0.111	0.053	0.103	0.027	0.107	0.056	0.029	0.000
#5	0.000	0.004	0.009	0.000	0.800	0.053	0.044	0.012	0.114	0.046	0.043	0.003
#6	0.000	0.005	0.000	0.027	0.029	0.738	0.026	0.028	0.062	0.036	0.002	0.002
#7	0.000	0.000	0.011	0.012	0.030	0.030	0.499	0.153	0.086	0.018	0.000	0.002
#8	0.004	0.073	0.000	0.013	0.000	0.000	0.170	0.737	0.069	0.051	0.004	0.000
#9	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.000	0.193	0.000	0.000	0.000
#10	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.046	0.676	0.405	0.000
#11	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.013	0.080	0.000	0.322	0.042
#12	0.000	0.000	0.007	0.000	0.000	0.000	0.002	0.001	0.037	0.025	0.165	0.923

Table 4: Estimated PDC value on γ band in epoch 1

4.3 Change of brain connectivity across epochs

We repeat the same procedure for the rest epochs and set the best VAR order separately. Figure 13 demonstrates the

AIC curves of the first 15 epochs, from which we can see some epochs reach the lowest AIC at $\hat{d} = 3$ and some of

them are $\hat{d} = 4$. Table 5 shows the distribution of \hat{d} across all 247 epochs. We fit $\text{VAR}(\hat{d})$ to each epoch and estimate the corresponding connectivity matrices by LASSLE method. Figure 14 shows the boxplots of ACF and PACF of residuals fitted from tetrode 7 across all 247 epochs, which is a strong evidence that the residuals from tetrode 7 are white noise. Same phenomenons are observed for residuals fitted from other tetrodes.

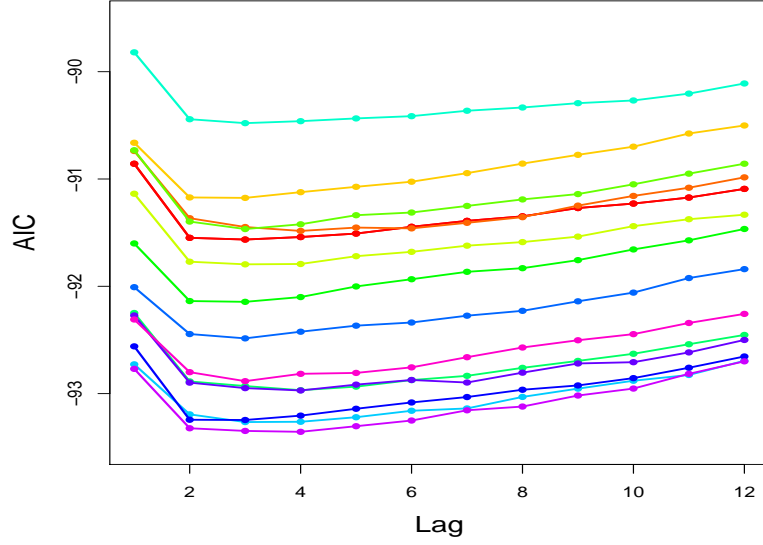


Figure 13: AIC of fitted VAR on first 15 epochs, lag order range: 1,2,...,12.

Selected lag order	2	3	4	5
Number of epochs	64	158	23	2

Table 5: Distribution of selected VAR lag order

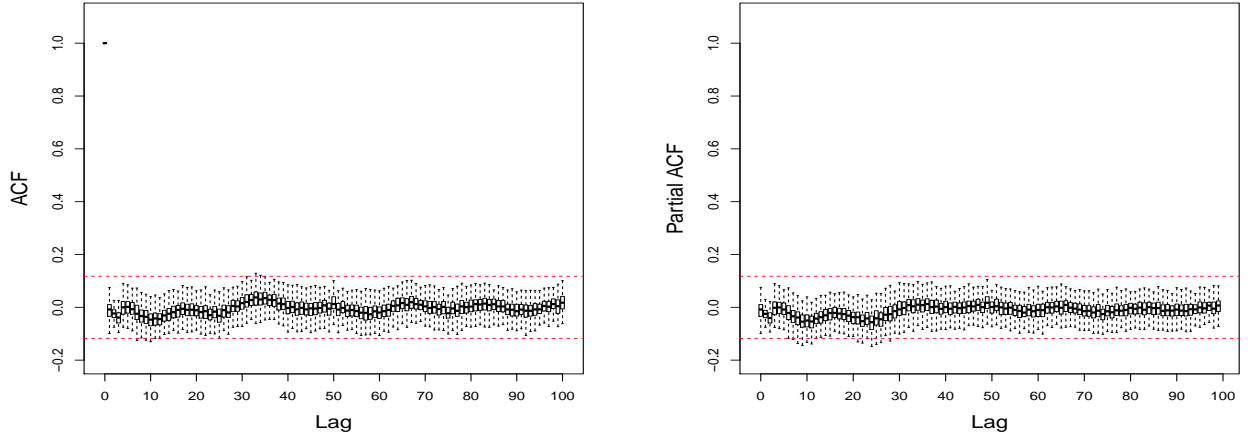


Figure 14: The boxplots of auto-correlation function (ACF) of residuals fitted from tetrode 7 across all epochs (left). The boxplots of partial auto-correlation function (PACF) of residuals fitted from tetrode 7 across all epochs (right).

After computing all the PDCs, we can obtain its 95% credible interval by summarizing from the empirical distribution. Here we are more interested in the variation of PDCs across all epochs. From Figure 15, we can clearly see that estimated PDC on γ band are quite stable on some tetrode pairs, e.g., tetrode 1 to tetrode 1. However, the estimates of PDC of some tetrode pairs are varying with epochs.

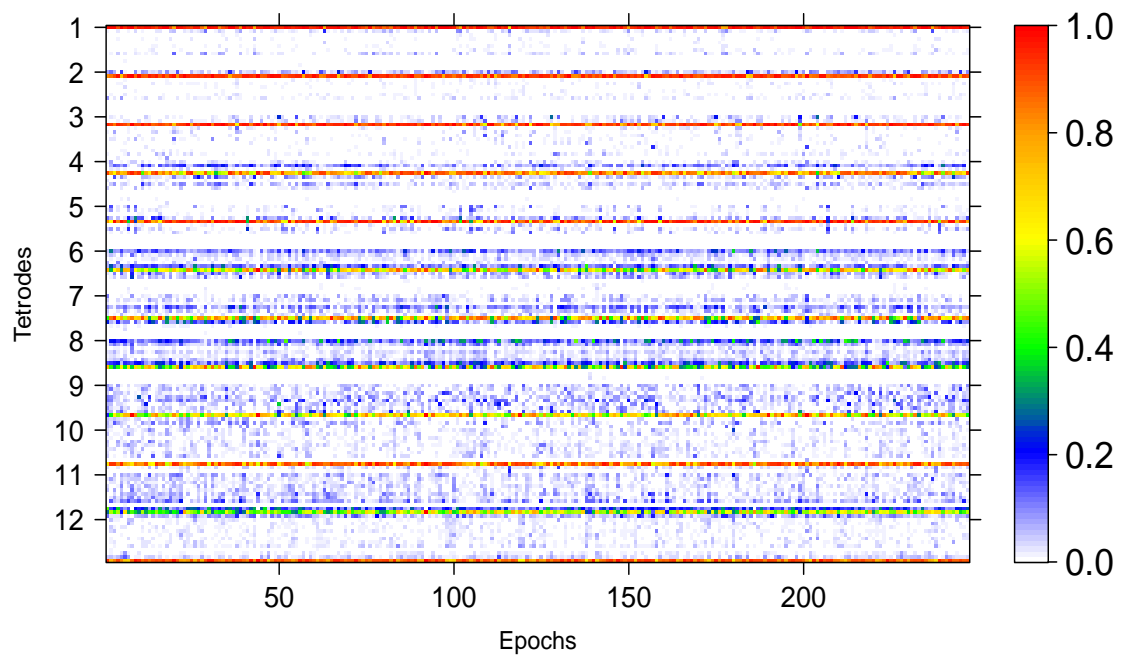


Figure 15: PDC on γ band across all epochs. X-axis is the index of epochs. At each epoch, PDC matrix (12×12) is converted to a column vector of $12 \times 12 = 144$ elements, where every 12 elements are the PDC values of one tetrode to all 12 tetrodes.

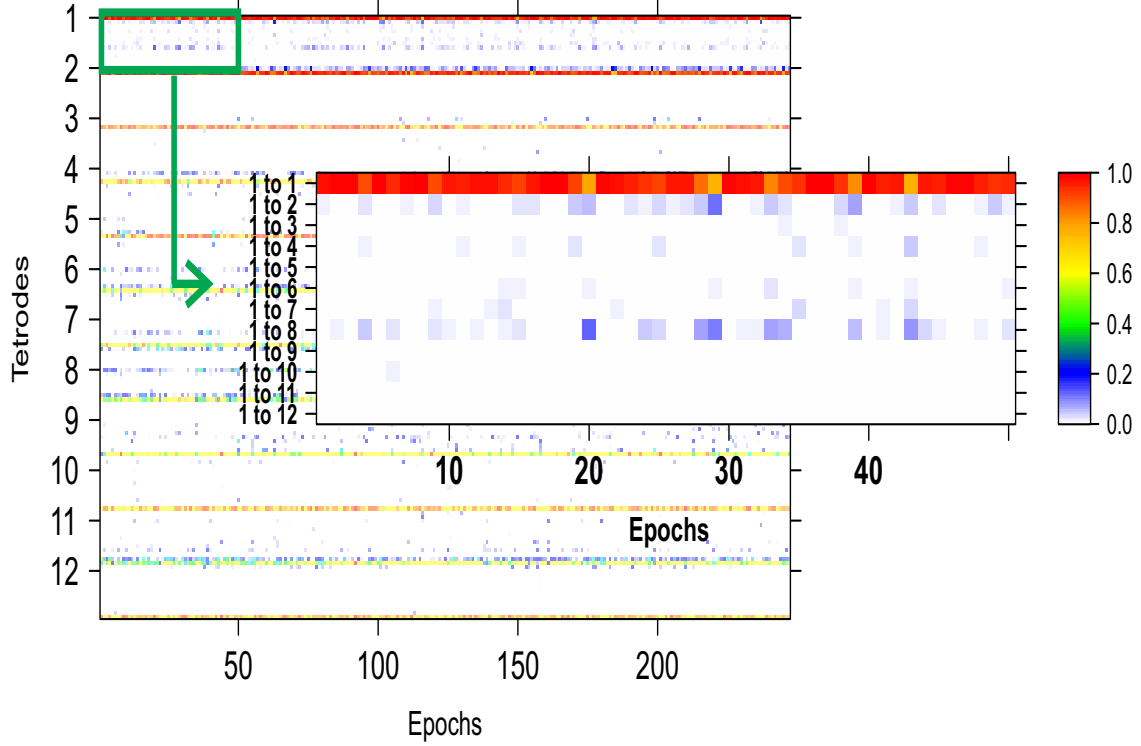


Figure 16: Illustration of Figure 15.

To compare the variation of estimated PDC on γ band between InSeq epochs and OutSeq epochs, we employ Kolmogorov-Smirnov (KS) test. The null hypothesis of KS test is that the empirical distribution of PDC from InSeq epochs and that of OutSeq epochs are identical. If the p-value is smaller than significant level $\alpha = 0.05$, then we reject H_0 and conclude that one distribution is different from the other one; otherwise we fail to reject H_0 and conclude that two distributions are the same. Here we use permutation to obtain the empirical distribution of KS test statistics. Since we want to preserve the inherent correlation across different epochs, we partition all 247 epochs to 50 groups where 5 consecutive epochs are within the same group (for the last group, we replicate epoch 246 and 247 to make 5 epochs). Then we randomly selected 5 groups (containing 25 epochs) from 50 groups as experimental OutSeq epochs, and let the rest to be experimental InSeq epochs. We compute the KS-statistic for this new Inseq and OutSeq grouping. We repeat this procedure for 10,000 times to obtain the empirical distribution of KS-statistics. At last we compute the proportion of permuted KS-statistics which have larger value than the real KS-statistic as the p-value.

Figure 17 and 18 demonstrate the empirical distributions of estimated PDC for all tetrodes given by Inseq epochs (blue curve) and Outseq epochs (red curve). Based on the p-values of KS test, there is strong evidence showing that the variation of auto-PDC of tetrode 1, 5, 7 and 8 are different between Inseq epochs and Outseq epochs. For

these tetrodes, the proportion of their current that can be explained by their own past is affected by the change of odor sequence. P-values also show that the variation of estimated PDC from tetrode 1, 5, 7 and 8 (sender) to some other tetrodes (receiver) are significantly different between InSeq epochs and OutSeq epochs, which means that the information flowing from these tetrodes to others are varying when the change of odor sequence occurs. This may suggest that the dorsal hippocampus area of tetrode 1, 5, 7 and 8 are more active when rats recognize the change of odor sequence. However, for the rest tetrodes the variation of their estimated auto-PDC are quite stable across InSeq epochs and OutSeq epochs.

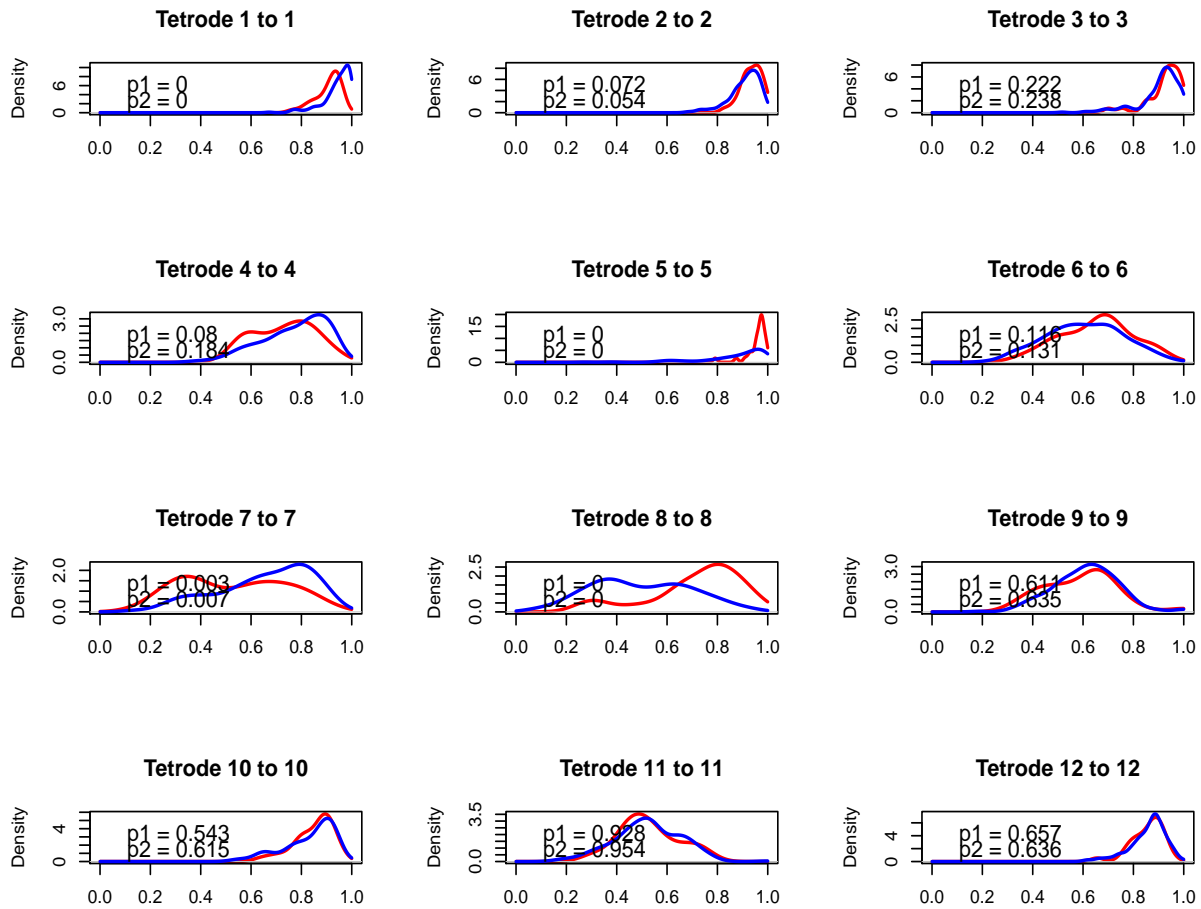


Figure 17: PDC density curves of diagonal entries across all 247 epochs. Blue one is the density curve of InSeq epochs. Red one is the density curve of OutSeq epochs only. P-value is calculated by Kolmogorov-Smirnov test. We reject the null hypothesis that two empirical distributions are the same when $p \leq 0.05$. P1 is the p-value given by R directly, while p2 is given by permutation.

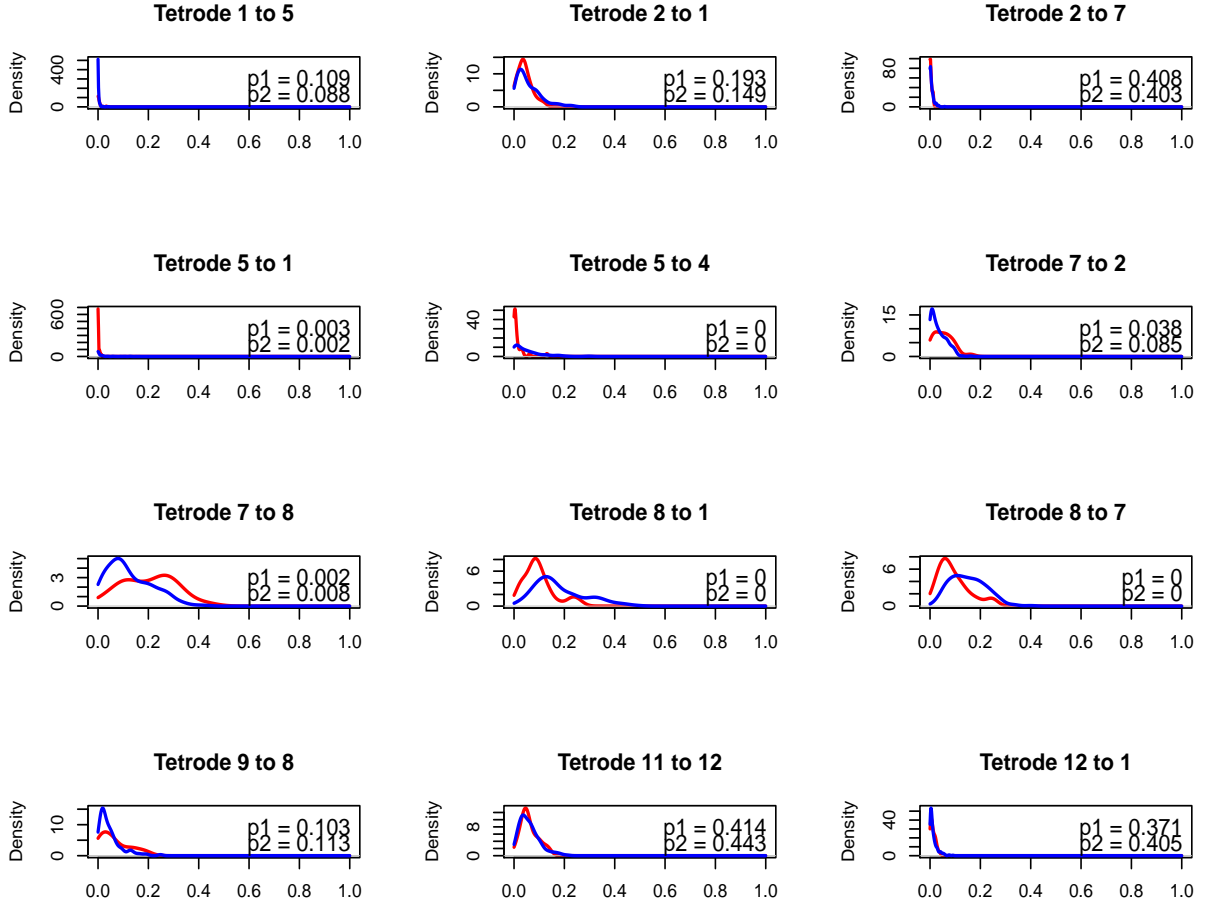


Figure 18: PDC density curves of some off-diagonal entries across all 247 epochs. Blue one is the density curve of InSeq epochs. Red one is the density curve of OutSeq epochs only. P-value is calculated by Kolmogorov-Smirnov test. We reject the null hypothesis that two imperial distributions are the same when $p \leq 0.05$. P1 is the p-value given by R directly, while p2 is given by permutation.

5 Conclusion

In this paper, we proposed hybrid LASSLE (LASSO+LSE) method to estimate the coefficients of sparse vector auto-regressive process characterizing the effective and directional connectivity for mul-tichannel brain physiological signals. This method use regularization to control for sparsity on the first stage and then use least squares to improve bias and mean-squared error of the estimator on the second stage. Compared to precious approach, the advantage of our method is that it is able to both indicate the most important effective connectivity and give a more accurate and precise estimate of the strength of connectivity. In addition we employ partial directed coherence to measure the directional connectivity between the channels. PDC is a directed frequency-specific measure that explains the extent

to which the present oscillatory activity in a sender channel influences the future oscillatory activity in a specific receiver channel relative all possible receivers in the network. Using the proposed modeling approach, we have achieved some insights on learning in a rat engaged in a non-spatial memory task.

References

- Timothy A Allen, Andrea M Morris, Aaron T Mattfeld, Craig EL Stark, and Norbert J Fortin. A sequence of events model of episodic memory shows parallels in rats and humans. *Hippocampus*, 24(10):1178–1188, 2014.
- Timothy A Allen, Daniel M Salz, Sam McKenzie, and Norbert J Fortin. Nonspatial sequence coding in cal neurons. *Journal of Neuroscience*, 36(5):1547–1563, 2016.
- Luiz A Baccalá and Koichi Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biological cybernetics*, 84(6):463–474, 2001.
- Luiz A Baccalá and Koichi Sameshima. Partial directed coherence. In *Methods in Brain Connectivity Inference through Multivariate Time Series Analysis*, pages 57–73. CRC Press, 2014.
- Mark Fiecas and Hernando Ombao. The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *The Annals of Applied Statistics*, pages 1102–1125, 2011.
- Mark Fiecas, Hernando Ombao, Crystal Linkletter, Wesley Thompson, and Jerome Sanes. Functional connectivity: Shrinkage estimation and randomization test. *NeuroImage*, 49(4):3005–3014, 2010.
- Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.
- Wenjiang J Fu. Penalized regressions: the bridge versus the lasso. *Journal of computational and graphical statistics*, 7(3):397–416, 1998.
- Fang Han and Han Liu. A direct estimation of high dimensional stationary vector autoregressions. *arXiv preprint arXiv:1307.0293*, 2013.
- Tim Hesterberg, Nam Hee Choi, Lukas Meier, Chris Fraley, et al. Least angle and l1 penalized regression: A review. *Statistics Surveys*, 2:61–93, 2008.
- Ventzislav Ivanov, Lutz Kilian, et al. A practitioner’s guide to lag order selection for var impulse response analysis. *Studies in Nonlinear Dynamics and Econometrics*, 9(1):1–34, 2005.

- Maciej Kamiński, Mingzhou Ding, Wilson A Truccolo, and Steven L Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological cybernetics*, 85(2):145–157, 2001.
- Hanzhong Liu, Bin Yu, et al. Asymptotic properties of lasso+ mls and lasso+ ridge in sparse high-dimensional linear regression. *Electronic Journal of Statistics*, 7:3124–3169, 2013.
- Julien Mairal and Bin Yu. Complexity analysis of the lasso regularization path. *arXiv preprint arXiv:1205.0079*, 2012.
- Hernando Ombao and Sébastien Van Bellegem. Evolutionary coherence of nonstationary signals. *IEEE Transactions on Signal Processing*, 56(6):2259–2266, 2008.
- CR Quirk, TA Allen, and NJ Fortin. Temporary inactivations of the hippocampus and prfrontal cortex impair memory for sequences of events. In *Soc Neurosci Abstr*, volume 39, 2013.
- Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- Peng Zhao and Bin Yu. On model selection consistency of lasso. *Journal of Machine learning research*, 7(Nov): 2541–2563, 2006.
- Peng Zhao, Guilherme Rocha, and Bin Yu. The composite absolute penalties family for grouped and hierarchical variable selection. *The Annals of Statistics*, pages 3468–3497, 2009.